International Journal for Uncertainty Quantification, 5 (3): 209–231 (2015)

SECOND-ORDER SENSITIVITY ANALYSIS OF PARAMETER ESTIMATION PROBLEMS

Ekaterina Kostina & Max Nattermann*

Institute for Apllied Mathematics, Heidelberg University, Im Neuenheimer Feld 293, 69120 Heidelberg, Germany

Original Manuscript Submitted: 10/15/2013; Final Draft Received: 01/21/2015

The use of model-based simulation to gain knowledge of unknown phenomena and processes behavior is a challenging task in many natural sciences. In order to get a full description of an underlying process, an important issue is to estimate unknown parameters from real but erroneous observations. Thus the whole system is affected by uncertainties and a sensitivity analysis is necessary. Usually one applies first-order sensitivity analysis and resulting linearized confidence regions to determine the statistical accuracy of the solution to parameter estimation problems. But especially in significantly nonlinear cases linearized regions may not be an adequate representation. In this paper, we suggest quadratic regions based on the second-order sensitivity analysis. The new region definition is based on a map that transforms the input uncertainties onto the parameter space. Furthermore, the approximation accuracy of the quadratic confidence regions is exemplary illustrated at two examples.

KEY WORDS: *uncertainty quantification, representing of uncertainty, inverse problems, parameter estimation, maximum likelihood, stochastic sensitivity analysis*

1. INTRODUCTION

In simulating real-world systems by using mathematical models, it is a challenging task to adapt the model to the process behavior by estimating unknown quantities. A common approach to identifying unknown model parameters is the minimization of the difference between the model response and real observations at certain times in a suitable norm. Due to inexact measuring methods, the measurements always contain defects, which lead to random influences and uncertainties in the computed solution.

Suppose that we have a set of real observations $\eta := (\eta_1, \ldots, \eta_{m_1})^T$ at corresponding times t_i, $i = 1, \ldots, m_1$. If the function h denotes the mathematical description of the measured system output, $\overline{x}$ denotes the vector consisting of the *true* parameter values and ε_i the corresponding measurement errors, then the measurements may be written as

$$\eta_i = h(t_i, \overline{x}) + \varepsilon_i,\ i = 1, \ldots, m_1.$$

It is a common assumption that the measurement errors follow a certain statistical distribution with zero means and known, or unknown, variances. Due to the fact that the measurements contain defects, the computed estimate x^* is affected by uncertainties. Hence, a sensitivity analysis is necessary to determine the statistical accuracy of x^*. In order to do this, the use of confidence regions is a common approach. Confidence regions are domains around the nominal parameter value, in such a way that the true values lie in this region with a certain probability.

In the case of only moderate nonlinear observation functions, without any restrictions on the parameter space, the procedure of quantifying uncertainties is well investigated. In order to minimize the computational costs (even in high-dimensional situations), frequently linearization techniques are used. A good overview of uncertainty quantification by using linear confidence regions can be found in Bard [1], Draper and Smith [2], Gallant [3], Pázman [4], and Seber and Wild [5].

*Correspond to Ekaterina Kostina, E-mail: kostina@uni-heidelberg.de

2152–5080/15/$35.00

Nowadays, mathematical modeling and simulation of time-dependent systems get more and more attention in many fields of application. Even in biological and chemical applications, the interest of getting knowledge about unknown behaviors and phenomena by using mathematical simulation is permanently increasing. This leads to the situation that mathematicians are confronted with a new kind of parameter estimation problems, which are often characterized by a very high nonlinearity. Because of the high complexity of the emerging problems, the common methods of quantifying errors are not longer adequate and there is a need for new investigations. To reinforce that linearized methods are not sufficient in many situations, we would like to refer to Donaldson and Schnabel [6], Rooney and Biegler [7], Schwaab et al. [8], Wiechert et al. [9], to name but a few.

The challenging requirements for the construction of a new confidence region are that the approximation accuracy should be as high as possible, and furthermore that the computations should be numerically tractable with low to moderate costs.

In literature, there are papers dealing with the construction of practical and suitable nonlinear confidence regions. Hamilton et al. [10] defined a confidence region by using a quadratic approximation to the solution locus and geometric concepts of sample space. Potocký and van Ban [11] developed confidence regions based on linear or quadratic approximations of curvature measures for nonlinear regression models. Dalai et al. [12] derived a method for the construction of confidence regions based on higher order statistics and an extension of the Leave-out Sign-Dominant Correlation Regions (LSCR) method. According to the authors, the resulting confidence regions are characterized by a guaranteed probability for any finite number of data samples. Modelhy and Marzouk [13] follow a Baysian inference approach and they quantify the influence of prior errors by the construction of map which transforms the randomness from the sample space onto the parameter space.

In this paper, we present a new definition of an approximation of confidence region based on a second-order sensitivity analysis. The new region is based on a quadratic map which transforms the random effects of the erroneous input data onto the parameter space. The construction of the map is based on functions that are basically available if one uses the Gauss-Newton method to solve the underlying parameter estimation problem.

In general, the definition of the new confidence region is not dependent on a certain error distribution. In this paper we concentrate on the situation with independent and normally distributed errors. Provided that the errors are independent and normally distributed, we get a maximum likelihood estimate of the parameters by minimizing the l_2-norm of the difference between model and data. We analyze the new region and present some bounds of the new confidence region. An important result is that there is a strong analogy between the quadratic approximation and the local convergence rate of the Gauss-Newton method.

Moreover, a quadratic approximation of the covariance matrix is presented, which leads to another tool for a higher order sensitivity analysis.

This paper is organized as follows. In Section 2 we give a brief introduction into parameter estimation problems. After discussing some basic formulations and conditions, we choose a generalized Gauss-Newton method to solve the problems. Section 3 presents a second-order representation of the estimated parameters depending on the prior errors. This representation serves as the basis of definition of the new confidence region. In Section 4 we discuss the confidence regions that are commonly used to perform a sensitivity analysis. The novel quadratic approximation of confidence regions is the topic of Section 5. After the definition we present and discuss bounds and other properties of the new region. In Section 6 we compute a quadratic approximation of the covariance matrix. Finally, in Section 7 we consider some numerical examples to demonstrate features of the quadratic approximation of the confidence region.

2. PARAMETER ESTIMATION PROBLEMS

We consider equality-constrained parameter estimation problems in general form

$$\begin{aligned} &\min_{x\in\mathbb{R}^{n_x}} \frac{1}{2}\left\|F_1(x)\right\|_2^2 \\ &\text{s.t. } F_2(x)=0, \end{aligned} \tag{1}$$

where the objective function consists of the weighted differences between erroneous measurements $\eta_1,\ldots,\eta_{m_1}$, and a corresponding measuring function $h:\mathbb{R}\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_{c_i}}\to\mathbb{R}$, at certain times, i.e.,

$$F_1(x) := \Sigma^{-1} \begin{pmatrix} \eta_1 - h(t_1, x, c_1) \\ \vdots \\ \eta_{m_1} - h(t_{m_1}, x, c_{m_1}) \end{pmatrix} \in \mathbb{R}^{m_1}. \tag{2}$$

The independent variables $t_i \in \mathbb{R}$ denote the measuring times, and the vector $x \in \mathbb{R}^{n_x}$ consists of unknown constants—the so-called parameters—which we want to estimate. By the vectors $c_i \in \mathbb{R}^{n_{c_i}}$ we denote the control variables of the system. By variations of the system controls, the experimenter is able to produce measurements under different conditions, which are more or less qualified for estimating parameters. Moreover, we assume that the measurement errors ε_i, $i = 1, \ldots, m_1$, are normally distributed with zero mean and the corresponding variances σ_i^2. The matrix Σ denotes the regular diagonal matrix $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_{m_1})$. Note that due to the assumptions of the error distribution, the solution of problem (1) delivers a maximum likelihood estimate.

In the constrained function $F_2 : \mathbb{R}^{n_x} \to \mathbb{R}^{m_2}$, the whole system behavior and given restrictions can be modeled. Frequently, F_2 consists of the discretized model equations, describing the process behavior and some potential conditions like boundary conditions, initial value conditions and the like.

Both functions $F_i : D \subset \mathbb{R}^{n_x} \to \mathbb{R}^{m_i}$ are assumed to be at least twice continuously differentiable, and the Jacobians are given by the matrices $J_i(x) := \partial F_i / \partial x \in \mathbb{R}^{m_i \times n_x}$, $i = 1, 2$. Furthermore, we assume that $m_2 < n_x$ and $m_1 + m_2 \geq n_x$. To ensure the solvability of problem (1), the Jacobians have to satisfy two regularity conditions on the domain D. The first condition is the *Constrained Qualification* (CQ), where we assume that $rank\ J_2(x) = m_2$. The second condition is the so-called *Positive Definiteness* (PD), where $rank\ J(x) = n_x$, with $J^T(x) := \left(J_1^T(x), J_2^T(x)\right)$. If both assumptions are fulfilled, we make sure that we have on the one hand enough information to estimate all unknown parameters (assumption PD), and on the other hand that it is guaranteed that there are no contradictions and no redundancies in the constraints (assumption CQ).

Basically, any suitable optimization method can be applied to solve problem (1). The method of our choice is the generalized Gauss-Newton method, as it was introduced by Bock [14] and Nocedal and Wright [15]. The advantages of this method are its good performance properties and the necessity of the first-order derivatives only. The algorithm starts with an initial guess x^0, and the solution is computed iteratively. A new iterate is updated by the rule

$$x^{k+1} = x^k + t^k \Delta x^k,\ k = 0, 1, 2, \ldots, \tag{3}$$

where the step size $t^k \in (0, 1]$ is determined by a suitable line search (see e.g., Nocedal and Wright [15]), and the increment $\Delta x^k \in \mathbb{R}^{n_x}$ is the solution of the linearized problem

$$\begin{aligned} &\min_{\Delta x^k} \frac{1}{2} \| F_1(x^k) + J_1(x^k) \Delta x^k \|_2^2 \\ &\text{s.t. } F_2(x^k) + J_2(x^k) \Delta x^k = 0. \end{aligned} \tag{4}$$

The algorithm stops if an appropriate termination condition is fulfilled, e.g., $\| \Delta x^k \| \leq tol$, where tol is a given tolerance. The main cost of one Gauss-Newton iteration is basically determined by solving the linearized system (4). According to the optimality conditions of the linearized problem in step k, the solution Δx^k, as well as the increment $\Delta \lambda^k$ of the Lagrange multiplier λ^k, may be written as

$$\begin{pmatrix} \Delta x^k \\ \Delta \lambda^k \end{pmatrix} = - \begin{pmatrix} J_1^T(x^k) J_1(x^k) & J_2^T(x^k) \\ J_2(x^k) & 0 \end{pmatrix}^{-1} \begin{pmatrix} J_1^T(x^k) & 0 \\ 0 & \mathbb{I} \end{pmatrix} \begin{pmatrix} F_1(x^k) \\ F_2(x^k) \end{pmatrix}.$$

Note that the regularity of the Karush-Kuhn-Tucker (KKT) matrix on the domain D can be easily shown by the conditions (CQ) and (PD), see [14]. By introducing the notations

$$J^+(x) := \begin{pmatrix} \mathbb{I} & 0 \end{pmatrix} \begin{pmatrix} J_1^T(x) J_1(x) & J_2^T(x) \\ J_2(x) & 0 \end{pmatrix}^{-1} \begin{pmatrix} J_1^T(x) & 0 \\ 0 & \mathbb{I} \end{pmatrix},\ F(x) := \begin{pmatrix} F_1(x) \\ F_2(x) \end{pmatrix} \tag{5}$$

the increment of a Gauss-Newton iteration can be written by the matrix-vector product $\Delta x^k = -J^+(x^k) F(x^k)$. The operator J^+ is a generalized inverse, but in general it is not a Moore-Penrose pseudo-inverse, since the axiom $(JJ^+)^T = JJ^+$ is not necessarily fulfilled. However, at least the relation $J^+(x)J(x) = \mathbb{I}_{n_x}$ holds.

The local convergence properties of the Gauss-Newton method will be discussed in the following theorem.

Theorem 1. *(Local Contraction Theorem) (Bock [14])*

Let $F^T := \left(F_1^T, F_2^T\right) \in \mathcal{C}^1(D, \mathbb{R}^{m_1+m_2})$, with the corresponding Jacobian $J^T = \left(J_1^T, J_2^T\right) = \partial F^T/\partial x$. Furthermore, we assume that J_1 and J_2 satisfy the regularity assumptions (CQ) and (PD) on D and J^+ denotes the generalized inverse according to (5). For all vectors $x \in D$ with $y := x + J^+(x)F(x)$, which implies that $\Delta x = y - x$ and $t \in (0, 1]$ we assume that:

- *there exists an $\omega < \infty$ such that*

$$\|J^+(y)\Big(J(x + t\Delta x) - J(x)\Big)\Delta x\| \leq \omega t\|\Delta x\|^2;$$

- *there exists a $\kappa(x) \leq \kappa < 1$ such that*

$$\|J^+(y)R(x)\| \leq \kappa(x)\|\Delta x\| \tag{6}$$

with the residual $R(x) = \Big(\mathbb{I} - J(x)J^+(x)\Big)F(x)$.

Further, we assume that the given initial value $x_0 \in D$ satisfies the relations

$$\delta_0 := \kappa + \frac{\omega}{2}\|\Delta x_0\| < 1,$$

$$\delta_k := \kappa + \frac{\omega}{2}\|\Delta x_k\|, \text{ where } \Delta x_k := -J^+(x_k)F(x_k)$$

and

$$D_0 := \left\{ z \in \mathbb{R}^n \mid \|z - x_0\| \leq \frac{\|\Delta x_0\|}{1 - \delta_0} \right\} \subseteq D.$$

Then the following hold:

1. *the full-step Gauss-Newton iterations are well-defined and remain in D; there exists an $x^* \in D$ with $x^k \to x^*$ for $k \to \infty$ such that $J^+(x^*)F(x^*) = 0$;*
2. *the convergence rate is linear with*

$$\|\Delta x^{k+1}\| \leq \left(\kappa + \frac{\omega}{2}\|\Delta x^k\|\right)\|\Delta x^k\|;$$

3. *an a priori estimate is given by*

$$\|x_{k+j} - x^*\| \leq \frac{\delta_0^{k+j}}{1 - \delta_0}\|\Delta x_0\|.$$

The proof of Theorem 1 follows lines of the Banach fixed point theorem and is explored thoroughly in [14]. According to Theorem 1 we have asymptotic linear convergence with rate κ. The Lipschitz constant ω is basically determined by $\|J^+\| \cdot \|dJ\|$ in some norm, where $dJ := \partial J/\partial x$ denotes the second derivative, and it can be seen as a measure for the curvature of the nonlinear function F. The significant meaning of κ is that only if $\kappa < 1$, a compatibility between the model and the measurements can be guaranteed, see Bock [14] and Bock et al. [16]. One main result of this paper is that these Lipschitz constants are also useful to determine the statistical accuracy of the estimate x^*. We will see that the bounds on the quadratic approximation of confidence regions are characterized by the constants ω and κ and a first-order approximation of the covariance matrix.

3. PARAMETER REPRESENTATION

In this section we present a first- and a second-order representation of the unknown parameter vector as a function of an error weight $\tau \in [0, 1]$. To this end, we consider a modified problem as

$$\begin{aligned} &\min_{x\in\mathbb{R}^{n_x}} \frac{1}{2}\|F_1(x,\tau)\|_2^2 \\ &\text{s.t. } F_2(x)=0. \end{aligned} \tag{7}$$

Taking (2) into account, the modified objective function F_1 is explicitly given by

$$F_1(x,\tau)=\Sigma^{-1}\begin{pmatrix} (h(t_1,\overline{x},c_1)+\tau\varepsilon_1) & - & h(t_1,x,c_1) \\ & \vdots & \\ (h(t_{m_1},\overline{x},c_{m_1})+\tau\varepsilon_{m_1}) & - & h(t_{m_1},x,c_{m_1}) \end{pmatrix},$$

where $\overline{x}$ denotes again the vector of the true parameter values. Note that, if $\tau=0$, the solution of problem (7) corresponds to $\overline{x}$, and if $\tau=1$, the original problem (1) and the modified problem (7) have the same solution. The optimality conditions of the modified problem together with the constraints are given by

$$\mathcal{F}(x,\lambda,\tau):=\begin{pmatrix} J_1^T(x,\tau)F_1(x,\tau)+J_2^T(x)\lambda \\ F_2(x) \end{pmatrix}=0, \tag{8}$$

where we introduce the function $\mathcal{F}:\mathbb{R}^{n_x}\times\mathbb{R}^{m_2}\times[0,1]\to\mathbb{R}^{n_x+m_2}$. Let the Jacobian of the function $\mathcal{F}$ be given by $\mathcal{J}(x,\lambda,\tau)=\partial\mathcal{F}/\partial(x,\lambda)$. For a more readable representation, we denote

$$\begin{aligned} &\mathcal{F}[\tau]:=\mathcal{F}(x(\tau),\lambda(\tau),\tau), \quad F_1[\tau]:=F_1(x(\tau),\tau), \quad F_2[\tau]:=F_2(x(\tau)), \\ &\mathcal{J}[\tau]:=\mathcal{J}(x(\tau),\lambda(\tau),\tau), \quad J_1[\tau]:=J_1(x(\tau),\tau), \quad J_2[\tau]:=J_2(x(\tau)). \end{aligned}$$

The following lemma gives a representation of the first derivative of the parameter vector as a function of τ. Using the derivative we get a first-order Taylor approximation of x.

Lemma 1. *(First-order representation)*
Let $x(0)=\overline{x}$ be the vector of the true parameters and assume that the Jacobians J_1 and J_2 satisfy the regularity assumptions (CQ) and (PD) in a neighborhood of $\overline{x}$. Then, for $\tau\in U_{\tau_0=0}$ the derivatives $\dot{x}(\tau)$ and $\dot{\lambda}(\tau)$ are uniquely defined by the system

$$\mathcal{J}[\tau]\begin{pmatrix}\dot{x}(\tau)\\ \dot{\lambda}(\tau)\end{pmatrix}=-\begin{pmatrix}J_1^T[\tau](\Sigma^{-1}\varepsilon)\\ 0\end{pmatrix}$$

and a first-order representation of the parameter vector is given by

$$\begin{aligned} x(\tau)&=x(0)+\tau\dot{x}(0)+\mathcal{O}(\tau^2)\\ &=\overline{x}+\tau J^+(\overline{x})\begin{pmatrix}\Sigma^{-1}\varepsilon\\ 0\end{pmatrix}+\mathcal{O}(\tau^2). \end{aligned}$$

Proof. The proof is given following Bock et al. [17] and is based on an application of the implicit function theorem. Let $x(\tau)$ be a solution of the modified problem, with $x(\tau)\to x(0)=\overline{x}$ if $\tau\to 0$. In the case of $\tau=0$, the true parameter vector $x(0)=\overline{x}$ is the solution of (7) and $F_1(\overline{x},0)=0$. Due to the assumption that we have the true model functions, the vector $\overline{x}$ is a feasible point, and thus the constraint $F_2(\overline{x})=0$ is fulfilled. Furthermore, it holds that $\lambda(0)=0$ because of the regularity assumption (CQ). The Jacobian $\mathcal{J}(x,\lambda,\tau):=\partial\mathcal{F}/\partial(x,\lambda)$ of the vector-valued function $\mathcal{F}(x,\lambda,\tau)$ is explicitly given by

$$\mathcal{J}[\tau]=\begin{pmatrix} J_1^T[\tau]J_1[\tau]+\frac{\partial J_1^T}{\partial x}[\tau](\mathbb{I}\otimes F_1[\tau])+\frac{\partial J_2^T}{\partial x}[\tau](\mathbb{I}\otimes\lambda) & J_2^T[\tau] \\ J_2[\tau] & 0 \end{pmatrix},$$

where

$$(\mathbb{I}\otimes F_1[\tau]):=\begin{pmatrix}F_1[\tau] & & 0\\ & \ddots & \\ 0 & & F_1[\tau]\end{pmatrix}\in\mathbb{R}^{(m_1\cdot n_x)\times n_x} \text{ and } (\mathbb{I}\otimes\lambda):=\begin{pmatrix}\lambda & & 0\\ & \ddots & \\ 0 & & \lambda\end{pmatrix}\in\mathbb{R}^{(m_2\cdot n_x)\times n_x}.$$

Evaluated at $\tau = 0$, the Jacobian reduces to

$$\mathcal{J}[0] = \begin{pmatrix} J_1^T[0]J_1[0] & J_2^T[0] \\ J_2[0] & 0 \end{pmatrix},$$

and according to the regularity conditions (CQ) and (PD) it is non-singular. Thus, at $\tau_0 = 0$, the assumptions of the implicit function theorem are fulfilled, and therefore there exist a τ_0-neighborhood U_{τ_0} and unique functions $x(\tau) : U_{\tau_0} \to \mathbb{R}^{n_x}$ and $\lambda(\tau) : U_{\tau_0} \to \mathbb{R}^{m_2}$ that satisfy the optimality condition (8) and the the initial conditions $x(0) = \overline{x}$ and $\lambda(0) = 0$. Moreover, there exists a neighborhood of $\tau_0 = 0$ such that for all τ in this neighborhood the derivatives $\dot{x}(\tau) := \partial x(\tau)/\partial\tau$ and $\dot{\lambda}(\tau) := \partial\lambda(\tau)/\partial\tau$ are the unique solution of the linear system

$$\mathcal{J}[\tau] \begin{pmatrix} \dot{x}(\tau) \\ \dot{\lambda}(\tau) \end{pmatrix} = -\frac{\partial \mathcal{F}[\tau]}{\partial \tau}, \text{ where } \frac{\partial \mathcal{F}[\tau]}{\partial \tau} = -\begin{pmatrix} J_1^T[\tau]\Sigma^{-1}\varepsilon \\ 0 \end{pmatrix}.$$

In particular, for $\tau = 0$ the derivative of the parameter vector is given by

$$\frac{\partial x(0)}{\partial \tau} = -J^+(\overline{x}) \begin{pmatrix} \Sigma^{-1}\varepsilon \\ 0 \end{pmatrix}.$$

The first-order representation follows directly by a Taylor expansion. □

Before we define the second-order representation of the parameter vector, in the following lemma we consider the derivative of the generalized inverse J^+.

Lemma 2. *Let $x \in \mathbb{R}^{n_x}$ be a feasible vector and assume that the Jacobians J_1 and J_2 satisfy the regularity assumptions (CQ) and (PD) in x. Then the derivative of the generalized inverse is given by*

$$\frac{\partial J^+(x)}{\partial x} = \mathcal{C}(x)\frac{\partial J^T(x)}{\partial x}\left(\begin{pmatrix} \mathbb{I} & 0 \\ 0 & 0 \end{pmatrix} - (J(x)J^+(x))^T\right) - J^+(x)\frac{\partial J(x)}{\partial x}J^+(x),$$

where

$$\mathcal{C}(x) := J^+(x)\begin{pmatrix} \mathbb{I}_{m_1} & 0 \\ 0 & 0_{m_2} \end{pmatrix} J^{+T}(x).$$

Proof. By definition it holds that

$$J^+(x) = \begin{pmatrix} \mathbb{I} & 0 \end{pmatrix} \mathcal{J}^{-1}(x) \begin{pmatrix} J_1^T(x) & 0 \\ 0 & \mathbb{I} \end{pmatrix}.$$

Considering

$$\frac{\partial J^+(x)}{\partial x} = -\begin{pmatrix} \mathbb{I} & 0 \end{pmatrix} \mathcal{J}^{-1}\frac{\partial \mathcal{J}(x)}{\partial x}\mathcal{J}^{-1}(x)\begin{pmatrix} J_1^T(x) & 0 \\ 0 & \mathbb{I} \end{pmatrix} + \begin{pmatrix} \mathbb{I} & 0 \end{pmatrix} \mathcal{J}^{-1}(x)\begin{pmatrix} \frac{\partial J_1^T(x)}{\partial x} & 0 \\ 0 & \mathbb{I} \end{pmatrix}$$

the lemma follows after a computation of $\partial\mathcal{J}(x)/\partial x$ and an appropriate simplification. □

Lemma 3. *Let $x(0) = \overline{x}$ be the vector of the true parameters and assume that the Jacobians J_1 and J_2 satisfy the regularity assumptions (CQ) and (PD) in a neighborhood of $\overline{x}$. Then, for $\tau \in U_{\tau_0=0}$ the following expansion holds*

$$x(\tau) = x(0) + \tau\dot{x}(0) + \frac{\tau^2}{2}\,\ddot{x}\,(0) + \mathcal{O}(\tau^3).$$

The derivatives are given by

$$\dot{x}(0) = -\begin{pmatrix}\mathbb{I} & 0\end{pmatrix}\mathcal{J}^{-1}[0]\begin{pmatrix}J_1^T[0] & 0\\ 0 & \mathbb{I}\end{pmatrix}\begin{pmatrix}\Sigma^{-1}\varepsilon\\ 0\end{pmatrix} = -J^+(\overline{x})\begin{pmatrix}\Sigma^{-1}\varepsilon\\ 0\end{pmatrix},$$
$$\ddot{x}\,(0) = -2\left(dJ^+(\overline{x})\left(\mathbb{I} - J(\overline{x})J^+(\overline{x})\right) + \frac{1}{2}J^+(\overline{x})(dJ(\overline{x}))(-J^+(\overline{x}))\right)\begin{pmatrix}\Sigma^{-1}\varepsilon\\ 0\end{pmatrix},$$

where

$$dJ^+(\overline{x}) = \sum_{i=0}^{n}\frac{\partial J^+(\overline{x})}{\partial x_i}\dot{x}_i(0) \text{ and } dJ(\overline{x}) = \sum_{i=0}^{n}\frac{\partial J(\overline{x})}{\partial x_i}\dot{x}_i(0).$$

Proof. In the proof of Lemma 1, we obtained that according to the implicit function theorem, the derivatives $\dot{x}(\tau)$ and $\dot{\lambda}(\tau)$ are defined on a domain U_{τ_0} and that they are continuously differentiable. Hence, we are able to compute the second derivative of x with respect to τ. For $\tau \in U_{\tau_0}$ we obtain

$$\begin{aligned}\frac{\partial^2 x(\tau)}{\partial\tau^2} &= -\frac{\partial}{\partial\tau}\left(\begin{pmatrix}\mathbb{I} & 0\end{pmatrix}\mathcal{J}^{-1}[\tau]\begin{pmatrix}J_1^T[\tau]\begin{pmatrix}\Sigma^{-1}\varepsilon\end{pmatrix}\\ 0\end{pmatrix}\right)\\ &= \begin{pmatrix}\mathbb{I} & 0\end{pmatrix}\mathcal{J}^{-1}[\tau]\frac{\partial\mathcal{J}[\tau]}{\partial\tau}\mathcal{J}^{-1}[\tau]\begin{pmatrix}J_1^T[\tau](\Sigma^{-1}\varepsilon)\\ 0\end{pmatrix} - (\mathbb{I},0)\,\mathcal{J}^{-1}[\tau]\begin{pmatrix}\frac{\partial J_1^T[\tau]}{\partial\tau}\left(\Sigma^{-1}\varepsilon\right)\\ 0\end{pmatrix}\end{aligned}$$

With Lemma 2 and by simplification it follows that

$$\ddot{x}\,(0) = -2\Big((dJ^+(\overline{x}))(\mathbb{I} - J(\overline{x})J^+(\overline{x})) + \frac{1}{2}J^+(\overline{x})(dJ(\overline{x}))(-J^+(\overline{x}))\Big)\begin{pmatrix}\Sigma^{-1}\varepsilon\\ 0\end{pmatrix}.$$

The second-order representation follows directly by a Taylor series and Lemma 1. □

Thus, a second-order representation of the parameter vector depends on the Jacobian J, the generalized inverse J^+, the second derivative dJ and the weighted measurement errors. Note that dJ is according to Lemma 2 the only second derivative we need. Especially in the context of optimum experimental design, where we also need the second derivative dJ, all matrix functions of the second-order representation in Lemma 3 are known.

There exists a remarkable relation between the introduced second-order parameter representation and the Lipschitz constants κ and ω, which are introduced in Theorem 1. In order to verify this, the following lemma gives an adequate estimate of the Lipschitz constant κ.

Lemma 4. *Let us assume that x is a feasible point and that the Jacobians J_1 and J_2 satisfy the regularity assumptions (CQ) and (PD). Furthermore, we introduce the following notations*

$$\begin{aligned}\lambda(x) &:= -\begin{pmatrix}0 & \mathbb{I}\end{pmatrix}\begin{pmatrix}J_1^T(x)J_1(x) & J_2^T(x)\\ J_2(x) & 0\end{pmatrix}^{-1}\begin{pmatrix}J_1^T(x) & 0\\ 0 & \mathbb{I}\end{pmatrix}F(x),\\ v(x) &:= F_1(x) - J_1(x)\mathcal{C}(x)J_1^T(x)F_1(x),\\ R(x) &:= F(x) - J(x)J^+(x)F(x),\\ \mathcal{C}(x) &:= J^+(x)\begin{pmatrix}\mathbb{I} & 0\\ 0 & 0\end{pmatrix}J^{+T}(x),\\ E(x) &:= \frac{\partial J_1^T(x)}{\partial x}\left(\mathbb{I}\otimes v(x)\right) + \frac{\partial J_2^T(x)}{\partial x}\left(\mathbb{I}\otimes\lambda(x)\right),\end{aligned}$$

where

$$\frac{\partial J_1^T(x)}{\partial x}\left(\mathbb{I}\otimes v(x)\right) := \left(\frac{\partial J_1^T(x)}{\partial x_1},\ldots,\frac{\partial J_1^T(x)}{\partial x_{n_x}}\right)\begin{pmatrix}v(x) & & 0\\ & \ddots & \\ 0 & & v(x)\end{pmatrix} \in \mathbb{R}^{m_1\times n_x}$$

and

$$\frac{\partial J_2^T(x)}{\partial x}\left(\mathbb{I}\otimes\lambda(x)\right) := \left(\frac{\partial J_2^T(x)}{\partial x_1},\ldots,\frac{\partial J_2^T(x)}{\partial x_{n_x}}\right)\begin{pmatrix}\lambda(x) & & 0\\ & \ddots & \\ 0 & & \lambda(x)\end{pmatrix}\in\mathbb{R}^{m_2\times n_x}.$$

Furthermore, we define

$$\tilde{\kappa}(x) := \left\|\mathcal{C}(x)E(x)\right\|.$$

1. *Then it holds that*

$$\begin{aligned}\left\|J^+(y)R(x)\right\| &= \left\|dJ^+(x)R(x)\right\| + \mathcal{O}(\|\Delta x\|^2)\\ &\leq \tilde{\kappa}\,\|\Delta x\| + \mathcal{O}(\|\Delta x\|^2),\end{aligned}$$

where $\Delta x := y - x = -J^+(x)F(x)$, with $x, y \in D \subseteq \mathbb{R}^{n_x}$ and the total derivative

$$dJ^+(x) := \sum_{i=1}^{n_x}\frac{\partial J^+(x)}{\partial x_i}\Delta x_i.$$

2. *In terms of κ as defined in Theorem 1 it holds that*

$$\tilde{\kappa} < 1 \Longrightarrow \kappa < 1.$$

3. *If x^* is an arbitrary point satisfying the KKT conditions of the equality-constrained parameter estimation problem, $E(x^*)$ reduces to*

$$E(x^*) = \frac{\partial J_1^T(x^*)}{\partial x}\left(\mathbb{I}\otimes F_1(x^*)\right).$$

Proof.

1. A first-order Taylor series of $J^+(y)$ around x yields

$$J^+(y) = J^+(x) + \sum_{i=1}^{n_x}\frac{\partial J^+(x)}{\partial x_i}\Delta x_i + \mathcal{O}(\|\Delta x\|^2),$$

and because of $J^+(x)R(x) = 0$ we get

$$\left\|J^+(y)R(x)\right\| = \left\|dJ^+(x)R(x)\right\| + \mathcal{O}(\|\Delta x\|^2),$$

where

$$dJ^+(x) := \sum_{i=1}^{n_x}\frac{\partial J^+(x)}{\partial x_i}\Delta x_i.$$

Recognizing Lemma 2, where the derivative of the generalized inverse is given, we obtain the following equation

$$\begin{aligned}dJ^+(x)R(x) &= \left(\mathcal{C}(x)\,dJ^T(x)\left(\begin{pmatrix}\mathbb{I} & 0\\ 0 & 0\end{pmatrix} - \left(J(x)J^+(x)\right)^T\right) - J^+(x)(dJ(x))J^+(x)\right)R(x)\\ &= \mathcal{C}(x)\,dJ^T(x)\left(\begin{pmatrix}\mathbb{I} & 0\\ 0 & 0\end{pmatrix} - \left(J(x)J^+(x)\right)^T\right)\left(\mathbb{I} - J(x)J^+(x)\right)\begin{pmatrix}F_1(x)\\ 0\end{pmatrix}\\ &= \mathcal{C}(x)\,dJ^T(x)\left(\mathbb{I} - \left(J(x)J^+(x)\right)^T\right)\begin{pmatrix}F_1(x)\\ 0\end{pmatrix},\end{aligned}$$

where we used

$$\left(\begin{pmatrix}\mathbb{I} & 0\\ 0 & 0\end{pmatrix} - (J(x)J^{+}(x))^{T}\right)\left(\mathbb{I} - J(x)J^{+}(x)\right)\begin{pmatrix}F_1(x)\\ 0\end{pmatrix} = \left(\mathbb{I} - (J(x)J^{+}(x))^{T}\right)\begin{pmatrix}F_1(x)\\ 0\end{pmatrix}.$$

Furthermore, with

$$\mathcal{J}^{-1}(x) := \begin{pmatrix}\mathcal{C}(x) & Z^{T}(x)\\ Z(x) & T(x)\end{pmatrix} := \begin{pmatrix}J_1^{T}(x)J_1(x) & J_2^{T}(x)\\ J_2(x) & 0\end{pmatrix}^{-1}$$

and $dJ^{T} = (dJ_1^{T}, dJ_2^{T})$ it holds that

$$\begin{aligned} dJ^{T}(x)\left(\mathbb{I} - (J(x)J^{+}(x))^{T}\right)\begin{pmatrix}F_1(x)\\ 0\end{pmatrix} &= dJ_1^{T}(x)\left(F_1(x) - \begin{pmatrix}\mathbb{I} & 0\end{pmatrix}\begin{pmatrix}J_1 & 0\\ 0 & \mathbb{I}\end{pmatrix}\mathcal{J}^{-1}(x)\begin{pmatrix}J_1^{T}(x)F_1(x)\\ 0\end{pmatrix}\right)\\ &\quad + dJ_2^{T}(x)\left(-\begin{pmatrix}0 & \mathbb{I}\end{pmatrix}\begin{pmatrix}J_1 & 0\\ 0 & \mathbb{I}\end{pmatrix}\mathcal{J}^{-1}(x)\begin{pmatrix}J_1^{T}(x)F_1(x)\\ 0\end{pmatrix}\right)\\ &= dJ_1^{T}(x)\left(F_1(x) - J_1(x)\mathcal{C}(x)J_1^{T}(x)F_1(x)\right) + dJ_2^{T}(x)\lambda(x)\\ &= dJ_1^{T}(x)v(x) + dJ_2^{T}(x)\lambda(x)\\ &= \left(\frac{\partial J_1^{T}(x)}{\partial x}\left(\mathbb{I}\otimes v(x)\right) + \frac{\partial J_2^{T}(x)}{\partial x}\left(\mathbb{I}\otimes\lambda(x)\right)\right)\Delta x. \end{aligned}$$

Therefore, we get

$$\left\|dJ^{+}(x)R(x)\right\| = \|\mathcal{C}(x)E(x)\Delta x\| \le \|\mathcal{C}(x)E(x)\|\,\|\Delta x\| \;= \tilde{\kappa}\|\Delta x\|.$$

2. If we reduce D in such a way that $\mathcal{O}(\|y-x\|^{2}) \le (1-\tilde{\kappa})\|y-x\|/2$ for all x, y, it holds that

$$\left\|J^{+}(y)R(x)\right\| = \tilde{\kappa}\|\Delta x\| + \mathcal{O}(\|y-x\|^{2}) \le \tilde{\kappa}\|\Delta x\| + \frac{1-\tilde{\kappa}}{2}\|y-x\| = \frac{1+\tilde{\kappa}}{2}\|y-x\| =: \overline{\kappa}\|\Delta x\|,$$

   where $\overline{\kappa} < 1$.

3. This follows since it holds that $J_1^{T}(x^{*})F_1(x^{*}) = 0$, if x^{*} satisfies the KKT conditions of the equality-constrained parameter estimation problem. □

Considering the second-order representation in any arbitrary norm, the triangular inequality yields

$$\frac{1}{2}\left\|\ddot{x}\,(0)\right\| \le \left\|dJ^{+}(\overline{x})(\mathbb{I} - J(\overline{x})J^{+}(\overline{x}))\begin{pmatrix}\Sigma^{-1}\varepsilon\\ 0\end{pmatrix}\right\| + \frac{1}{2}\left\|J^{+}(\overline{x})(dJ(\overline{x}))(-J^{+}(\overline{x}))\begin{pmatrix}\Sigma^{-1}\varepsilon\\ 0\end{pmatrix}\right\|. \tag{9}$$

The first expression of the right-hand side can be interpreted using Lemma 4. According to this lemma, it holds that

$$\left\|dJ^{+}(\overline{x})(\mathbb{I} - J(\overline{x})J^{+}(\overline{x}))\begin{pmatrix}\Sigma^{-1}\varepsilon\\ 0\end{pmatrix}\right\| = \left\|dJ^{+}(\overline{x})R(\overline{x})\right\| \le \tilde{\kappa}(\overline{x})\|\Delta x\|,$$

where

$$R(\overline{x}) := (\mathbb{I} - J(\overline{x})J^{+}(\overline{x}))\begin{pmatrix}\Sigma^{-1}\varepsilon\\ 0\end{pmatrix}.$$

Note that $\Sigma^{-1}\varepsilon = F_1(\overline{x})$. The second expression of (9) can be estimated by

$$\begin{aligned}\frac{1}{2}\left\|J^+(\overline{x})(dJ(\overline{x}))(-J^+(\overline{x}))\begin{pmatrix}\Sigma^{-1}\varepsilon\\0\end{pmatrix}\right\| &\leq \frac{1}{2}\left\|J^+(\overline{x})(dJ(\overline{x}))\right\|\|\Delta x\|\\ &\leq \frac{1}{2}\left\|J^+(\overline{x})\frac{\partial J(\overline{x})}{\partial x}\right\|\|\Delta x\|^2\\ &\leq \frac{1}{2}\left\|J^+(\overline{x})\right\|\cdot\left\|\frac{\partial J(\overline{x})}{\partial x}\right\|\|\Delta x\|^2\\ &=: \frac{1}{2}\tilde{\omega}(\overline{x})\|\Delta x\|^2.\end{aligned}$$

Obviously, we do not know the true parameter vector $\overline{x}$ in practice. We assume that the computed solution of the parameter estimation problem x^* is indeed a parameter estimate, i.e. it is a continuous deformation of the true parameter values, as a function of measurement errors, in a compact region. In this region we can further reasonably assume that the constants κ and ω are bounded. Indeed, ω is a weighted Lipschitz constant on the Jacobians and κ is a Lipschitz constant on the generalized inverse weighted by the least squares residuals. Moreover, the constant κ needs to be smaller than one for parameter estimation problems to be well-posed also in nonlinear case. Hence, in the absence of any other information we approximate the constant $\tilde{\kappa}(\overline{x})$ by $\tilde{\kappa}(x^*)$ and the constant $\tilde{\omega}(\overline{x})$ by $\tilde{\omega}(x^*)$ which is defined in a similar way to ω of the *Local Contraction Theorem* 1 by

$$\tilde{\omega}(x^*) := \left\|J^+(x^*)\right\|\cdot\left\|\frac{\partial J(x^*)}{\partial x}\right\|.$$

Thus, we get the following bound for the second derivative of the parameter values:

$$\frac{1}{2}\left\|\ddot{x}\ (0)\right\| \lesssim \left(\tilde{\kappa}(x^*) + \frac{\tilde{\omega}(x^*)}{2}\left\|\Delta x\right\|\right)\left\|\Delta x\right\|.$$

The bound depends on Lipschitz constants $\tilde{\kappa}$ as well as on $\tilde{\omega}$ with a squared weight $\|\Delta x\|^2$.

4. CONFIDENCE REGIONS

The existence of erroneous input data leads to uncertainties in the computed solution x^* and therefore a sensitivity analysis is necessary. In order to get information about the accuracy of an estimate, we need to know how uncertainties in the observation space are propagated into the parameter space.

One approach of quantifying the quality of an estimated parameter vector is the already mentioned confidence regions. The idea of confidence regions is to define a domain $\mathcal{D} \subseteq \mathbb{R}^n$ surrounding the nominal parameter value x^*, in such a way that the true parameter vector $\overline{x}$ lies in this region with a certain probability $(1-\alpha)$. Obviously, confidence regions depend on the observations η—and on their uncertainties—as a part of the input data of the underlying parameter estimation problem. In addition the size of confidence regions is determined by the so-called confidence level $(1-\alpha)$, where $0 < \alpha < 1$. Obviously, the smaller the value of α, the bigger is the confidence region. However, $\mathcal{D}(\eta, \alpha)$ is a confidence region, if the equality

$$\mathbb{P}\left(\overline{x} \in \mathcal{D}(\eta, \alpha)\right) = 1 - \alpha, \tag{10}$$

or at least the inequality

$$\mathbb{P}\left(\overline{x} \in \mathcal{D}(\eta, \alpha)\right) \geq 1 - \alpha \tag{11}$$

holds, [4]. Basically, there are several possibilities to construct a confidence region, but some further requirements are preferable. On the one hand, confidence regions should be numerically well tractable. This means that the computation should be easy, fast and especially not error-prone. On the other hand, the confidence region should be as accurate as possible, which means as small as possible having regard to (10) or (11).

Before we define a new quadratic approximation of confidence region, we consider some common confidence regions for the estimate x^*.

Likelihood ratio confidence regions. If we consider an unconstrained parameter estimation problem—where we are confronted with the task

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\left\|F_1(x)\right\|_2^2, \tag{12}$$

using the notations from (2)—a confidence region for an estimate x^* is given by

$$\mathcal{DU}_{lr}(\alpha) = \left\{x \in D \mid \|F_1(x)\|_2^2 - \|F_1(x^*)\|_2^2 \leq \gamma^2_{n_x}(\alpha)\right\}, \tag{13}$$

see e.g., [2, 4, 18]. In case the standard deviations σ_i are known, $\gamma^2_{n_x}(\alpha)$ denotes the $(1-\alpha)$-quantile of the χ^2-distribution. If the values σ_i are unknown, we define $\gamma^2_{n_x}(\alpha) := s^2 \cdot n_x \cdot F_{n_x,m_1-n_x}$, where F_{n_x,m_1-n_x} is the $(1-\alpha)$-quantile of the F-distribution and

$$s^2 := \frac{\left\|F_1(x^2)\right\|^2}{m_1 - n_x}. \tag{14}$$

An asymptotically justification of the confidence region definition (13) can be derived by a simple likelihood ratio test, if we test a hypothesis x^* against another hypothesis $x \neq x^*$. This test results in a log-proportional expression like $\|F_1(x)\|_2^2 - \|F_1(x^*)\|_2^2$, and if under assumption that the measurement errors are independent and normally distributed, we get asymptotically (13), see Pázman [4].

Bock [14] adapted the nonlinear confidence region to constrained parameter estimation problems. If we consider problem (1), with an estimate x^*, a confidence region is given by

$$\mathcal{D}_{lr}(\alpha) := \left\{x \in D \mid F_2(x) = 0, \|F_1(x)\|_2^2 - \|F_1(x^*)\|_2^2 \leq \gamma^2_{\overline{m}}(\alpha)\right\}, \tag{15}$$

where $\overline{m} := n_x - m_2$ denotes the degrees of freedom.

The good approximation properties of likelihood ratio confidence regions must be paid by a huge complexity and very high computational costs, especially in significantly nonlinear cases. The computation of (15), or at least of some appropriate bounds, requires the solution of a nonlinear equation with $\overline{m}$ degrees of freedom. Due to this, likelihood ratio confidence regions are not practicable in many applications, see e.g. Vanrolleghem and Keesman [19].

Linearized confidence regions. To counteract the high computational costs of likelihood ratio confidence regions, a common approach is to apply linearization techniques, see [2, 4, 5, 14]. By a first-order Taylor expansion of $\mathcal{D}_{lr}(\alpha)$ we obtain the linearized confidence region

$$\begin{aligned}\mathcal{D}_{lin}(\alpha) := \{x \in D \mid\ & F_2(x^*) + J_2(x^*)(x - x^*) = 0,\\ & \|F_1(x^*) + J_1(x^*)(x - x^*)\|_2^2 - \|F_1(x^*)\|_2^2 \leq \gamma^2_{\overline{m}}(\alpha)\},\end{aligned}$$

and due to the optimality conditions for x^* we can rewrite $\mathcal{D}_{lin}(\alpha)$ as

$$\mathcal{D}_{lin}(\alpha) = \{x \in D \mid J_2(x^*)(x - x^*) = 0, \|J_1(x^*)(x - x^*)\|_2^2 \leq \gamma^2_{\overline{m}}(\alpha)\}, \tag{16}$$

and in the unconstrained case

$$\mathcal{DU}_{lin}(\alpha) := \{x \in D \mid \|J_1(x^*)(x - x^*)\|_2^2 \leq \gamma^2_{n_x}(\alpha)\}, \tag{17}$$

see [14]. The meanings of $\gamma^2_{\overline{m}}(\alpha)$ and $\gamma^2_{n_x}(\alpha)$, respectively, remain unaffected by the linearizations and are described above.

The shape of the linearized confidence region is characterized by an ellipsoid and it is very cheap to compute this region. In case of a linear observation function h, these regions are optimal in the sense that they have a minimal volume with a confidence level exactly equal to $(1-\alpha)$. However, this holds only for the linear case, and in literature many nonlinear applications can be found, where the elliptical regions are a poor approximation, see e.g., [6–9].

Operator-based linear confidence regions. Another approach to performing a sensitivity analysis is based on the linear operator J^+.

According to Lemma 1, a first-order Taylor series of x as a function of the error weight is given by

$$x(\tau) = \overline{x} + \tau \frac{\partial x(0)}{\partial \tau} + \mathcal{O}(\tau^2) = \overline{x} - \tau J^+(\overline{x}) \begin{pmatrix} -\Sigma^{-1}\varepsilon \\ 0 \end{pmatrix} + \mathcal{O}(\tau^2). \tag{18}$$

Remember that the measurement errors are assumed to be independent and normally distributed with zero mean and variances σ_i^2, $i = 1, \ldots, m_1$. Hence, the first-order approximation of $x(\tau)$ is normally distributed, too, with expected value $\mathbb{E}\left(x(\tau)\right) = \overline{x}$ and covariance matrix

$$\overline{\mathcal{C}} = J^+(\overline{x}) \begin{pmatrix} \mathbb{I}_{m_1} & 0 \\ 0 & 0_{m_2} \end{pmatrix} J^{+T}(\overline{x}).$$

In practice, we use the estimate x^* instead of the unknown vector $\overline{x}$. The use of x^* is justified by the expectation that the solution of the generalized Gauss-Newton method is a good approximation of the true values $\overline{x}$. Thus, in our further considerations we use the covariance matrix

$$\mathcal{C} := J^+(x^*) \begin{pmatrix} \mathbb{I}_{m_1} & 0 \\ 0 & 0_{m_2} \end{pmatrix} J^{+T}(x^*). \tag{19}$$

This matrix is symmetric, positive semi-definite, and has $rank(\mathcal{C}) = \overline{m} = n_x - m_2$.

Another linearized confidence region can be defined with the help of the first-order, error-depending representation of x. If x^* is the solution of problem (1) and the Jacobians J_1 and J_2 satisfy the regularity assumptions (CQ) and (PD) in x^*, a linear confidence region is given by

$$\overline{\mathcal{D}}_{lin}(\alpha) := \left\{ x^* + \Delta x \mid \Delta x = -J^+(x^*) \begin{pmatrix} \eta \\ 0 \end{pmatrix}, \|\eta\|_2^2 \leq \gamma_{\overline{m}}^2(\alpha) \right\}. \tag{20}$$

Following Bock et al. [17], we can show that $\mathcal{D}_{lin}(\alpha) = \overline{\mathcal{D}}_{lin}(\alpha)$. Hence the properties of (16) can be adapted to the region (20). Furthermore, the following lemma shows that the exact bounds on the region (20) are related to the diagonal elements of the covariance matrix (19).

Lemma 5. *Let x^* be a solution of problem (1) and assume that the Jacobians J_1 and J_2 satisfy the regularity assumptions (CQ) and (PD) in x^*. Then, $\mathcal{D}_{lin}(\alpha)$ is contained in a minimal box defined by the cross product of the confidence intervals,*

$$\mathcal{D}_{lin}(\alpha) \subset \bigtimes_{i=1}^{n_x} [x_i^* - \theta_i, x_i^* + \theta_i],$$

where $\theta_i = \left(C_{ii}\gamma_{\overline{m}}^2(\alpha)\right)^{1/2}$. The values C_{ii} denote the diagonal elements of the covariance matrix $\mathcal{C}$. Furthermore, it holds

$$\max_{x \in \mathcal{D}_{lin}(\alpha)} |x_i - x_i^*| = \theta_i, \quad i = 1, \ldots, n.$$

The proof of this lemma is e.g., given in [17]. According to Lemma 5, it is sufficient to compute the diagonal elements of the covariance matrix to perform a first-order sensitivity analysis.

5. A QUADRATIC APPROXIMATION OF CONFIDENCE REGIONS

To pursue the idea of the linearized region (20), we suggest a quadratic approximation of confidence regions based on a second-order sensitivity analysis. Considering Lemma 3, a quadratic approximation of confidence regions is defined as follows.

Definition 5.1. Let x^* be the solution of problem (1) and assume that the Jacobians J_1 and J_2 satisfy the regularity assumptions (CQ) and (PD) in x^*. Then, a quadratic approximation of a confidence region is defined by

$$\begin{aligned}\overline{\mathcal{D}}_{quad}(\alpha) := \Bigg\{ & x^* + \Delta x + \frac{1}{2}\overline{\Delta x} \mid \Delta x = -J^+ \begin{pmatrix} \eta \\ 0 \end{pmatrix}, \\ & \overline{\Delta x} = -2 \left(dJ^+(\mathbb{I} - JJ^+) - \frac{1}{2} J^+ (dJ) J^+ \right) \begin{pmatrix} \eta \\ 0 \end{pmatrix}, \ \|\eta\|_2^2 \leq \gamma^2_{\overline{m}}(\alpha) \Bigg\},\end{aligned}$$

where all the functions are evaluated at x^*, and the total derivatives are given by

$$\begin{aligned} dJ(x^*) &= \frac{\partial J(x^*)}{\partial x} (\Delta x \otimes \mathbb{I}) = \sum_{i=1}^{n} \frac{\partial J(x^*)}{\partial x_i} \left(e_i^T \Delta x \right), \\ dJ^+(x^*) &= \frac{\partial J^+(x^*)}{\partial x} (\Delta x \otimes \mathbb{I}) = \sum_{i=1}^{n} \frac{\partial J^+(x^*)}{\partial x_i} \left(e_i^T \Delta x \right). \end{aligned}$$

We want to remark that the derivative of the Jacobian J is the only second derivative that is needed to compute the quadratic approximation of confidence regions. The derivative of the generalized inverse J^+ is explicitly given in Lemma 2. For further information of the computation of matrix derivatives we refer to Magnus and Neudecker [20].

For the sake of completeness, we want to note that the new confidence region is of course also usable in the unconstrained case (12). Here, a quadratic approximation of confidence regions is given by

$$\begin{aligned}\overline{\mathcal{DU}}_{quad}(\alpha) := \Bigg\{ & x^* + \Delta x + \frac{1}{2}\overline{\Delta x} \mid \Delta x = -J^+ \eta, \\ & \overline{\Delta x} = -2 \left(dJ^+(\mathbb{I} - J_1 J^+) - \frac{1}{2} J^+ (dJ_1) J^+ \right) \eta, \ \|\eta\|_2^2 \leq \gamma^2_{n_x}(\alpha) \Bigg\},\end{aligned} \tag{21}$$

where $J^+ = (J_1^T J_1)^{-1} J_1^T$ is a Moore-Penrose pseudo-inverse and all the functions are evaluated at x^*.

Lemma 6. *Let x^* be a solution of problem (1) and assume that the Jacobians J_1 and J_2 satisfy the regularity assumptions (CQ) and (PD) in x^*. Then*

$$\max_{\|\eta\|_2^2 \leq \gamma^2_{\overline{m}}(\alpha)} \frac{1}{2}\overline{\Delta x}_i = \mu^* \gamma^2_{\overline{m}}(\alpha), \qquad i = 1, \ldots, n_x,$$

where μ^* is the maximum eigenvalue of the symmetric matrix $-1/2 \sum_{i=1}^{n_x} \left(r_{j,i} c_i^T + c_i r_{j,i}^T \right)$ with

$$\begin{aligned} c_i^T &:= e_i^T (-J^+) \begin{pmatrix} \mathbb{I} \\ 0 \end{pmatrix} \\ r_{j,i}^T &:= -e_j^T \left(\frac{\partial J^+}{\partial x_i} \left(\mathbb{I} - JJ^+ \right) \begin{pmatrix} \mathbb{I} \\ 0 \end{pmatrix} + \frac{1}{2} J^+ \frac{\partial J}{\partial x_i} (-J^+) \begin{pmatrix} \mathbb{I} \\ 0 \end{pmatrix} \right), \end{aligned}$$

and all the functions are assumed to be evaluated at x^*.

Proof. It holds that

$$\begin{aligned} \frac{1}{2}\overline{\Delta x} &= - \left(dJ^+ \left(\mathbb{I} - JJ^+ \right) \begin{pmatrix} \eta \\ 0 \end{pmatrix} + \frac{1}{2} J^+ (dJ)(-J^+) \begin{pmatrix} \eta \\ 0 \end{pmatrix} \right) \\ &= - \sum_{i=1}^{n_x} \left(\frac{\partial J^+}{\partial x_i} \left(\mathbb{I} - JJ^+ \right) \begin{pmatrix} \mathbb{I} \\ 0 \end{pmatrix} + \frac{1}{2} J^+ \frac{\partial J}{\partial x_i} (-J^+) \begin{pmatrix} \mathbb{I} \\ 0 \end{pmatrix} \right) \left(c_i^T \eta \right) \end{aligned}$$

and hence

$$\frac{1}{2}\overline{\Delta x}_j = \sum_{i=1}^{n_x} -e_j^T \left(\frac{\partial J^+}{\partial x_i} \left(\mathbb{I} - JJ^+ \right) \begin{pmatrix} \mathbb{I} \\ 0 \end{pmatrix} + \frac{1}{2} J^+ \frac{\partial J}{\partial x_i} (-J^+) \begin{pmatrix} \mathbb{I} \\ 0 \end{pmatrix} \right) \left(c_i^T \eta \right) \eta = \sum_{i=1}^{n_x} \left(r_{j,i}^T \eta \right) \left(c_i^T \eta \right).$$

In order to find the maximum $\max\limits_{\|\eta\|_2^2 \leq \gamma_{\overline{m}}^2(\alpha)} (1/2)\overline{\Delta x}_i$, we consider the Lagrangian

$$\mathcal{L}(\eta, \mu) = -\sum_{i=1}^{n_x} \left(r_{j,i}^T \eta \right) \left(c_i^T \eta \right) - \mu \left(\gamma_{\overline{m}}^2(\alpha) - \eta^T \eta \right)$$

and the necessary optimality condition

$$0 = \frac{\partial \mathcal{L}(\eta, \mu)}{\partial \eta} = \left(-\sum_{i=1}^{n_x} \left(r_{j,i} c_i^T + c_i r_{j,i}^T \right) + 2\mu \mathbb{I} \right) \eta.$$

Thus we get that the matrix $-\sum\limits_{i=1}^{n_x} \left(r_{j,i} c_i^T + c_i r_{j,i}^T \right)$ has an eigenvalue 2μ with the corresponding eigenvector η. Furthermore, the necessary optimality condition yields

$$0 = \left(-\sum_{i=1}^{n_x} \left(r_{j,i} c_i^T + c_i r_{j,i}^T \right) + 2\mu \mathbb{I} \right) \eta \iff 0 = -\sum_{i=1}^{n_x} \left(\eta^T r_{j,i} \right) \left(c_i^T \eta \right) + \left(\eta^T c_i \right) \left(r_{j,i}^T \eta \right) + 2\mu \eta^T \eta$$

$$\iff 2\mu \gamma_{\overline{m}}^2(\alpha) = 2 \sum_{i=1}^{n_x} \left(\eta^T r_{j,i} \right) \left(c_i^T \eta \right)$$

and we have

$$\max_{\|\eta\|_2^2 \leq \gamma_{\overline{m}}^2(\alpha)} \frac{1}{2} \overline{\Delta x}_i = \mu^* \gamma_{\overline{m}}^2(\alpha),$$

where μ^* is the maximum eigenvalue of the matrix

$$-\sum_{i=1}^{n_x} \left(r_{j,i} c_i^T + c_i r_{j,i}^T \right).$$

□

In the following lemma we introduce bounds on the quadratic approximation of confidence region for each component of the parameter vector.

Lemma 7. *Let x^* be a solution of problem (1) and assume that the Jacobians J_1 and J_2 satisfy the regularity assumptions (CQ) and (PD) in x^*. Then $\overline{\mathcal{D}}_{quad}(\alpha)$ is contained in a box defined by the cross product of the confidence intervals,*

$$\overline{\mathcal{D}}_{quad}(\alpha) \subset \bigtimes_{i=1}^{n_x} [x_i^* - \theta_i, x_i^* + \theta_i],$$

where

$$\theta_i = \sqrt{C_{ii}} \gamma_{\overline{m}}(\alpha) + \sum_{k=1}^{n_x} \sqrt{\left(\widetilde{J}_k \widetilde{J}_k^T \right)_{ii} C_{kk}} \cdot \gamma_{\overline{m}}^2(\alpha),$$

with

$$\widetilde{J}_k := \frac{\partial J^+(x^*)}{\partial x_k} (\mathbb{I} - J(x^*) J^+(x^*)) - \frac{1}{2} J^+(x^*) \frac{\partial J(x^*)}{\partial x_k} J^+(x^*).$$

The constants C_{ii} denote the diagonal elements of the covariance matrix $\mathcal{C}$, $i = 1, \ldots, n_x$. Here all the matrix functions are evaluated at the solution x^*.

Proof. The lemma follows from the following relations

$$\begin{aligned}\left|\left(\Delta x+\frac{1}{2}\overline{\Delta x}\right)_i\right| &= \left|e_i^T J^+\begin{pmatrix}\eta\\0\end{pmatrix}+e_i^T\left(dJ^+(\mathbb{I}-JJ^+)-\frac{1}{2}J^+dJJ^+\right)\begin{pmatrix}\eta\\0\end{pmatrix}\right|\\ &\le \|e_i^T J^+\|_2\cdot\|\eta\|_2+\left\|e_i^T\left(dJ^+(\mathbb{I}-JJ^+)-\frac{1}{2}J^+dJJ^+\right)\right\|_2\cdot\|\eta\|_2\\ &\le \sqrt{C_{ii}}\cdot\gamma_{\overline{m}}(\alpha)+\sum_{k=1}^{n_x}\left\|e_i^T\widetilde{J}_k\right\|_2|e_k^T J^+\eta|\cdot\gamma_{\overline{m}}(\alpha)\\ &\le \sqrt{C_{ii}}\cdot\gamma_{\overline{m}}(\alpha)+\sum_{k=1}^{n_x}\sqrt{\left(\widetilde{J}_k\widetilde{J}_k^T\right)_{ii}C_{kk}}\cdot\gamma_{\overline{m}}^2(\alpha),\end{aligned}$$

where all the functions are evaluated at x^*. □

The next lemma gives a further estimation of the new confidence region by using the two Lipschitz constants ω and κ of J^+ and J, respectively.

Lemma 8. *Under the assumptions of Lemma 7 the following inequality holds*

$$\left\|\Delta x+\frac{1}{2}\overline{\Delta x}\right\|_2\le\theta+\left(\tilde{\kappa}(x^*)+\frac{1}{2}\tilde{\omega}(x^*)\theta\right)\theta.$$

The value of θ *is defined by* $\theta := \sqrt{trace(\mathcal{C}(x^*))\gamma_{\overline{m}}^2(\alpha)}$, *where* $\mathcal{C}(x^*)$ *denotes the linear approximation of the covariance matrix, and* $\tilde{\kappa}(x^*)$ *and* $\tilde{\omega}(x^*)$ *are given by*

$$\begin{aligned}\tilde{\kappa}(x^*) &:= \|\mathcal{C}(x^*)E(x^*)\|,\\ \tilde{\omega}(x^*) &:= \left\|J^+(x^*)\frac{\partial J(x^*)}{\partial x}\right\|,\end{aligned}$$

where $E(x^*)=(\partial J_1^T(x^*)/\partial x)(\mathbb{I}\otimes F_1(x^*))$.

Proof. The lemma follows from Lemma 4 and (9). □

Let us again consider the interpretations of the Lipschitz constants κ and ω, following the *Local Contraction Theorem* 1. According to Lemma 8 we may conclude that the new confidence regions depend on the one hand on the nonlinearity of the model function ω, and on the other hand on κ, i.e. on the compatibility between the model and the real observations.

6. A QUADRATIC APPROXIMATION OF THE COVARIANCE MATRIX

In Lemma 5 we have seen that the linear confidence regions have a direct relation to the diagonal elements of the linear approximation of the covariance matrix. In particular, if C_{ii} denotes the ith diagonal element of the covariance matrix approximation (19), we get linear approximations of confidence intervals by

$$\left[x^*-\sqrt{C_{ii}\gamma_{\overline{m}}^2(\alpha)},\ x^*+\sqrt{C_{ii}\gamma_{\overline{m}}^2(\alpha)}\right], \tag{22}$$

$i=1,\dots,n_x$. In this section, we compute a quadratic approximation of the covariance matrix, to get another tool for a higher order sensitivity analysis, by replacing in (22) the diagonal elements of the linear covariance by the diagonal elements of the quadratic approximation of the covariance matrix.

According to Lemma 3, and taking the estimate x^* as a good approximation of the true parameter vector $\overline{x}$, we get up to the second order that

$$x(\varepsilon) := x(\tau = 1) = x^* - J^+ \begin{pmatrix} -\Sigma^{-1}\varepsilon \\ 0 \end{pmatrix} - \left[(dJ^+)\Big(\mathbb{I} - JJ^+\Big) - \frac{1}{2} J^+ (dJ) J^+ \right] \begin{pmatrix} -\Sigma^{-1}\varepsilon \\ 0 \end{pmatrix},$$

where all the functions are evaluated at x^*. For the further considerations, let the inverse of the KKT-matrix explicitly be given by

$$\begin{pmatrix} X & Y^T \\ Y & Z \end{pmatrix} := \begin{pmatrix} J_1^T(x^*) J_1(x^*) & J_2^T(x^*) \\ J_2(x^*) & 0 \end{pmatrix}^{-1}. \tag{23}$$

If we take into account that the expected value is linear and that the measurement errors ε_i are normally distributed with the zero mean and the variances σ_i^2, the expected value of $x(\varepsilon)$ is given by

$$\begin{aligned} \mathbb{E}(x(\varepsilon)) &= \mathbb{E}(x^*) + J^+(x^*)\mathbb{E}\begin{pmatrix} \Sigma^{-1}\varepsilon \\ 0 \end{pmatrix} + \sum_{i=1}^{n_x} \widetilde{J}_i(x^*)\mathbb{E}\left[\left(e_i^T J^+(x^*) \begin{pmatrix} \Sigma^{-1}\varepsilon \\ 0 \end{pmatrix} \right) \begin{pmatrix} \Sigma^{-1}\varepsilon \\ 0 \end{pmatrix} \right] \\ &= x^* + \frac{1}{\sigma^2} \sum_{i=1}^{n_x} \widetilde{J}_i(x^*)\, \mathbb{E}\left[\left(e_i^T X J_1^T(x^*) \Sigma^{-1}\varepsilon \right) \begin{pmatrix} \Sigma^{-1}\varepsilon \\ 0 \end{pmatrix} \right] = x^* + \sum_{i=1}^{n_x} \widetilde{J}_i(x^*) \begin{pmatrix} J_1(x^*) X \\ 0 \end{pmatrix} e_i, \end{aligned} \tag{24}$$

where we introduced the notation

$$\widetilde{J}_i(x^*) := \frac{\partial J^+(x^*)}{\partial x_i}\Big(\mathbb{I} - J(x^*)J^+(x^*)\Big) - \frac{1}{2} J^+(x^*) \frac{\partial J(x^*)}{\partial x_i} J^+(x^*).$$

Now, we compute a quadratic approximation of the covariance matrix:

$$\begin{aligned} \mathcal{C}_2 :&= \mathbb{E}\Big[\Big(x(\varepsilon) - \mathbb{E}(x(\varepsilon)) \Big) \Big(x(\varepsilon) - \mathbb{E}(x(\varepsilon)) \Big)^T \Big] = \mathbb{E}\Big(x(\varepsilon) x(\varepsilon)^T \Big) - \mathbb{E}(x(\varepsilon))\mathbb{E}(x(\varepsilon))^T \\ &= X + \sum_{i=1}^{n_x} \sum_{k=1}^{n_x} \tilde{J}_i\, \mathbb{E}\left(\left(e_i^T X J_1^T \Sigma^{-1}\varepsilon \right) \begin{pmatrix} \Sigma^{-1}\varepsilon \\ 0 \end{pmatrix} \Big(\Sigma^{-1}\varepsilon^T, 0^T \Big) \Big(e_k^T X J_1^T \Sigma^{-1}\varepsilon \Big) \right) \widetilde{J}_k^T \\ &- \sum_{i=1}^{n_x} \sum_{k=1}^{n_x} \widetilde{J}_i \begin{pmatrix} J_1 X e_i \\ 0 \end{pmatrix} \Big(e_k^T X J_1^T \ ,\ 0^T \Big) \widetilde{J}_k^T = X + \sum_{i=1}^{n_x} \sum_{k=1}^{n_x} \left(\widetilde{J}_i\, x_{ik}\, \widetilde{J}_k^T + \widetilde{J}_i \begin{pmatrix} J_1 X e_i e_k^T X J_1^T & 0 \\ 0 & 0 \end{pmatrix} \widetilde{J}_k^T \right). \end{aligned} \tag{25}$$

Here all the functions are evaluated at x^* and $x_{ik} = e_i^T X e_k$. Important results from probability theory used here are that the nth power of an independent random variable is also independent and that the moment of order 4, of a normal distributed random variable with zero mean and variance σ^2, is $3\sigma^4$. Note that the matrix X from (23) is equal to the linear approximation of the covariance matrix (19).

7. NUMERICAL EXAMPLES

In this section, we want to show some comparative illustrations of the different confidence regions to get an idea of their shapes and their approximation accuracies.

Example 7.1. As a first example, we consider the biochemical oxygen demand (BOD) of stream water. The experimental data are taken from Marske [21], where also the setup of the experiments is described. According to Bates and Watts [22] the corresponding observation function is

$$h(t; x_1, x_2) = x_1 \cdot (1 - \exp(-t \cdot x_2)).$$

The variable t denotes the time (in days) and the two unknown model parameters x_1 and x_2 have to be estimated using 6 observations $(7.322416; 13.85557; 12.27182; 15.58857; 16.62757; 22.03010)$ at time points $(1; 2; 3; 4; 5; 7)$ [21]. There are no equality constraints in this parameter estimation problem.

An application of the *Gauss-Newton* method yields the optimal parameter values $x_1 = 19.1426$ and $x_2 = 0.5311$ with the corresponding linear covariance matrix

$$\mathcal{C} = \begin{pmatrix} 0.95876 & -0.066527 \\ -0.066527 & 0.0063474 \end{pmatrix}.$$

A comparison of the different confidence regions is given in Figs. 1 and 2. The solid lines illustrate the likelihood ratio confidence regions (13), the dotted lines illustrate the linear confidence regions (17), and the gray areas illustrate the quadratic approximations of the confidence regions (21). Obviously, the quadratic approximations of the confidence regions are more precise approximations of the likelihood ratio confidence regions than the linearized regions. In Table 1 a comparison of different confidence intervals with the probability levels $1 - \alpha = 0.995$ and $1 - \alpha = 0.95$ is illustrated. The intervals below the columns $\mathcal{D}_{lin}(\alpha)$, $\mathcal{D}_{quad}(\alpha)$ and $\mathcal{D}_{lr}(\alpha)$ are the exact bounds of the corresponding confidence regions. The values of $x_i^* - \tilde{\theta}_i$ and $x_i^* + \tilde{\theta}_i$, belonging to the confidence intervals of the

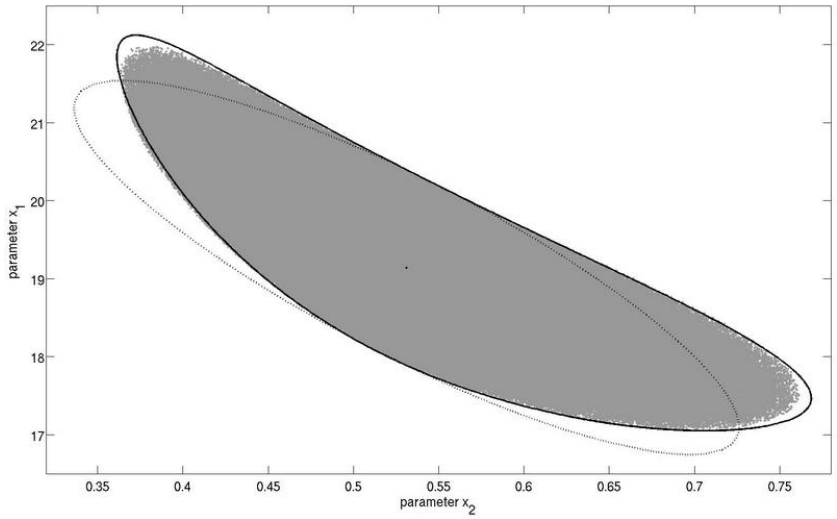

FIG. 1: Conf. reg. with $1 - \alpha = 0.95$.

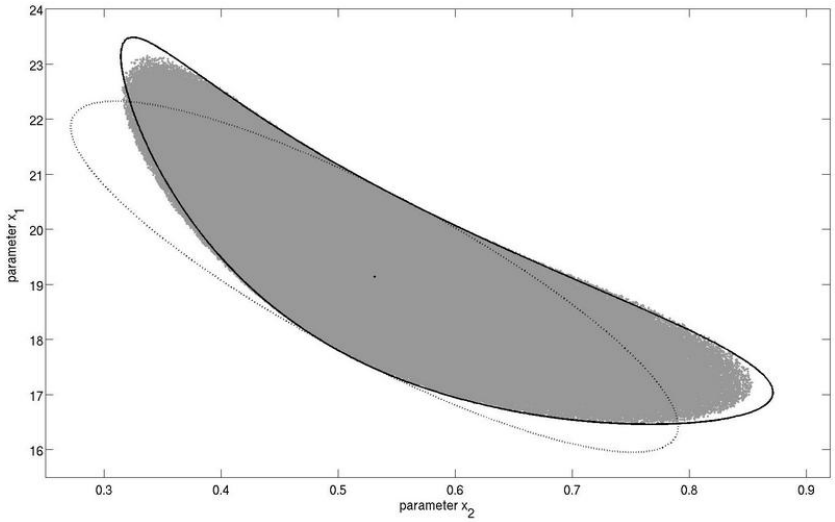

FIG. 2: Conf. reg. with $1 - \alpha = 0.995$.

TABLE 1: Confidence intervals for $1-\alpha = 0.95$ and $1-\alpha = 0.995$.

$1-\alpha$		$\mathcal{D}_{lin}(\alpha)$	$\mathcal{D}_{quad}(\alpha)$	$\mathcal{D}_{lr}(\alpha)$	$[x_i^* - \tilde{\theta}_i, x_i^* + \tilde{\theta}_i]$
0.95	x_1	$[16.746, 21.539]$	$[17.062, 21.983]$	$[17.053, 22.121]$	$[16.589, 21.695]$
	x_2	$[0.3360, 0.7261]$	$[0.3632, 0.7621]$	$[0.3613, 0.7683]$	$[0.3262, 0.7358]$
0.995	x_1	$[15.955, 22.330]$	$[16.475, 23.149]$	$[16.466, 23.484]$	$[15.595, 22.689]$
	x_2	$[0.2717, 0.7904]$	$[0.3153, 0.8544]$	$[0.3145, 0.8712]$	$[0.2491, 0.8130]$

last column, are computed by using the second-order approximation of the covariance matrix. More precisely, it holds $\tilde{\theta}_i = \sqrt{\mathcal{C}_{2,ii}\gamma_2^2(\alpha)}$, where $\mathcal{C}_{2,ii}$ is the ith diagonal element of the quadratic approximation of the covariance matrix (25). Here, it is remarkable that there is a strong similarity between the intervals of the likelihood ratio regions and the quadratic approximations of confidence regions. In particular, only these intervals are not symmetric around the estimate x^*.

Example 7.2. The second example deals with the energy y that is radiated from a carbon filament lamp per cm^2, depending on the temperature t. It is taken from Daniel and Wood [23] and Keeping [24], where you can find a more thorough treatment of the matter. The observation function is given by

$$h(t; x_1, x_2) = x_1 t^{x_2}$$

with the two unknown constants x_1 and x_2. The data contains 6 observations $(2.138; 3.421; 3.597; 4.340; 4.882; 5.660)$ corresponding to the values $(1.309; 1.471; 1.490; 1.565; 1.611; 1.680)$ of the absolute temperature of the filament in thousands of degrees K. There are no equality constraints in this parameter estimation problem.

The optimal parameter vector is $x^* = (0.7689, 3.86)^T$ with the linear approximation of the covariance matrix

$$\mathcal{C} = \begin{pmatrix} 0.30967 & -0.86808 \\ -0.86808 & 2.479 \end{pmatrix}.$$

A comparison of the different confidence regions is given in Figs. 3 and 4. As in the first example, the solid lines illustrate the likelihood ratio confidence regions (13), the dotted lines illustrate the linear confidence regions (17), and the gray areas illustrate the quadratic approximations of the confidence regions (21). Here it can also be seen that the quadratic approximations of the confidence regions are more precise approximations of the likelihood ratio confidence

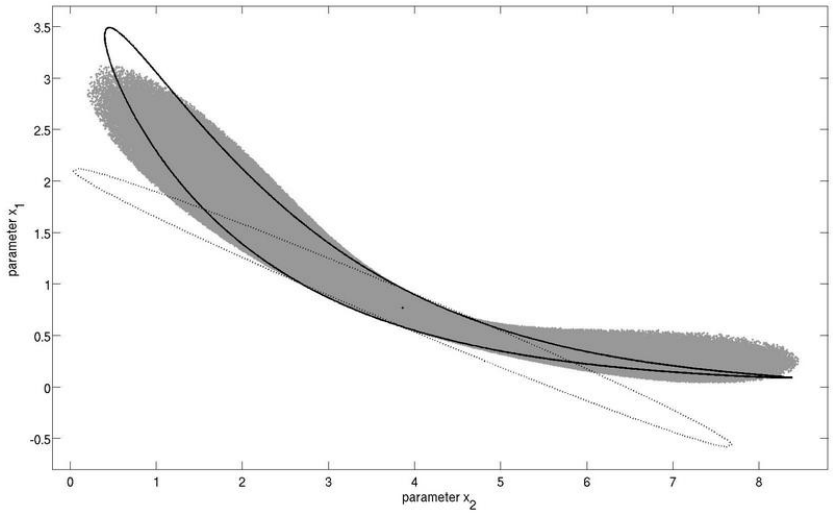

FIG. 3: Conf. reg. with $1-\alpha = 0.95$.

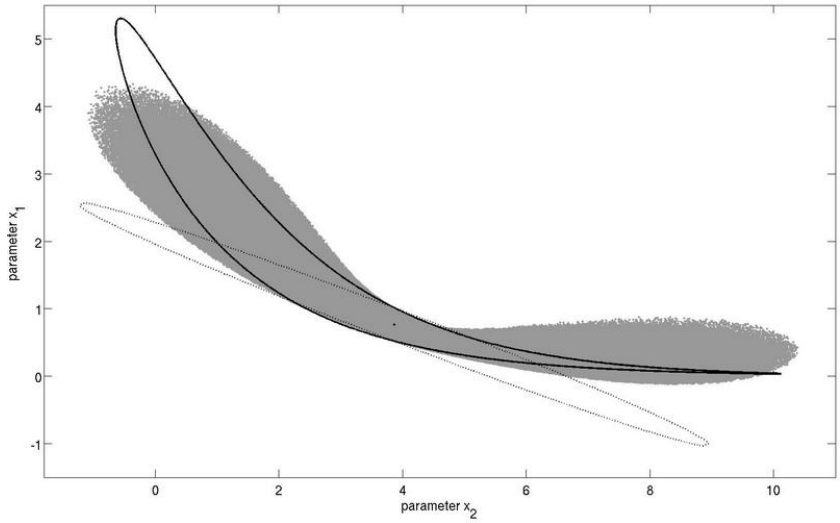

FIG. 4: Conf. reg. with $1 - \alpha = 0.995$.

regions than the linearized regions. A comparison of different confidence intervals with the probability levels $1 - \alpha = 0.995$ and $1 - \alpha = 0.95$ is illustrated in Table 2. The intervals below the columns $\mathcal{D}_{lin}(\alpha)$, $\mathcal{D}_{quad}(\alpha)$, and $\mathcal{D}_{lr}(\alpha)$ are the exact bounds of the corresponding confidence regions, and $\tilde{\theta}_i = \sqrt{\mathcal{C}_{2,ii}\gamma_2^2(\alpha)}$, where $\mathcal{C}_{2,ii}$ are the diagonal elements of the second-order approximation of the covariance matrix (25). Again, the non-symmetric intervals of the quadratic confidence regions are more precise approximations of the intervals belonging to the likelihood ratio regions.

Example 7.3. Further, we have performed various numerical experiments with parameter estimation problem of the second example using synthetic data. The observations for equidistantly distributed measurement points $t_i = 1.3 + (i-1)h, i = 1, \ldots, M$ at the temperature interval $T = [1.3, 1.79]$ have been simulated using "true" parameter values $\bar{x} = (0.7689, 3.8604)^T$. The observations have been corrupted by random measurement errors ε_i with zero mean and variances σ.

The results of the experiments are presented in Tables 3 and 4.

Figures 5–8 shows confidence regions for different measurement errors in the experiments with 5 observations. Confidence regions for measurement error with variances $\sigma = 1$ in the experiments with different numbers of observations are presented in Fig. 9–12.

The results of the experiments are in a very good agreement with the asymptotic behaviour to be expected with the increasing number of observations. The same can be said for the experiments with changing error variances.

Let us point out here that the above numerical investigations are not meant to suggest computation of quadratic approximations of confidence regions based on sampling since this would be computationally expensive in higher dimensions of parameter or measurement spaces. Rather the numerical experiments are to demonstrate significant distortion of confidence regions in the nonlinear case compared to linear approximations, as indicated by the Lipschitz constants ω and κ.

TABLE 2: Confidence intervals for $1 - \alpha = 0.95$ and $1 - \alpha = 0.995$

$1-\alpha$		$\mathcal{D}_{lin}(\alpha)$	$\mathcal{D}_{quad}(\alpha)$	$\mathcal{D}_{lr}(\alpha)$	$[x_i^* - \tilde{\theta}_i, x_i^* + \tilde{\theta}_i]$
0.95	x_1	$[-0.5932, 2.130]$	$[0.0448, 3.124]$	$[0.0894, 3.488]$	$[-1.226, 2.764]$
	x_2	$[0.00661, 7.714]$	$[0.2074, 8.453]$	$[0.4042, 8.384]$	$[-0.373, 8.094]$
0.995	x_1	$[-1.0426, 2.580]$	$[-0.116, 4.348]$	$[0.0391, 5.302]$	$[-2.382, 3.920]$
	x_2	$[-1.2649, 8.985]$	$[-1.095, 10.36]$	$[-0.638, 10.11]$	$[-2.129, 9.850]$

TABLE 3: Confidence intervals for different measurement errors and 5 observations

σ		x^*	$\mathcal{D}_{lin}(\alpha)$	$\mathcal{D}_{quad}(\alpha)$	$\mathcal{D}_{lr}(\alpha)$	$[x_i^* - \tilde{\theta}_i, x_i^* + \tilde{\theta}_i]$
0.1	x_1	0.83202	$[0.68257, 0.98147]$	$[0.69433, 0.98629]$	$[0.69235, 0.99194]$	$[0.68177, 0.98227]$
	x_2	3.6784	$[3.293, 4.0638]$	$[3.3019, 4.0649]$	$[3.2986, 4.0699]$	$[3.2926, 4.0642]$
0.3	x_1	0.66362	$[0.28086, 1.0464]$	$[0.36756, 1.1205]$	$[0.35946, 1.1304]$	$[0.26058, 1.0667]$
	x_2	4.1209	$[2.8963, 5.3455]$	$[2.9653, 5.3624]$	$[2.966, 5.397]$	$[2.8817, 5.3601]$
0.5	x_1	0.53103	$[-0.00071547, 1.0628]$	$[0.17966, 1.2564]$	$[0.17497, 1.272]$	$[-0.081103, 1.1432]$
	x_2	4.6263	$[2.5231, 6.7295]$	$[2.6642, 6.8808]$	$[2.7299, 6.8918]$	$[2.4504, 6.8022]$
1	x_1	0.29479	$[-0.35413, 0.94371]$	$[-0.020553, 1.455]$	$[0.01738, 1.6946]$	$[-0.7439, 1.3335]$
	x_2	5.974	$[1.46, 10.488]$	$[1.9498, 11.5596]$	$[2.1655, 11.6059]$	$[0.8017, 11.1463]$
2	x_1	0.56327	$[-2.0373, 3.1638]$	$[-0.52592, 7.6656]$	$[0.011393, 12.6247]$	$[-6.7818, 7.9084]$
	x_2	4.068	$[-5.7463, 13.8823]$	$[-6.369, 19.7902]$	$[-4.4652, 11.8672]$	$[-11.5786, 19.7146]$

TABLE 4: Confidence intervals for different number of observations and measurement error variance $\sigma = 1$

M, h		x^*	$\mathcal{D}_{lin}(\alpha)$	$\mathcal{D}_{quad}(\alpha)$	$\mathcal{D}_{lr}(\alpha)$	$[x_i^* - \tilde{\theta}_i, x_i^* + \tilde{\theta}_i]$
$5, 0.1$	x_1	0.29479	$[-0.35413, 0.94371]$	$[-0.020553, 1.455]$	$[0.01738, 1.6946]$	$[-0.7439, 1.3335]$
	x_2	5.974	$[1.46, 10.488]$	$[1.9498, 11.5596]$	$[2.1655, 11.6059]$	$[0.8017, 11.1463]$
$10, 0.05$	x_1	0.74559	$[-0.58062, 2.0718]$	$[0.042734, 3.2004]$	$[0.072778, 3.4968]$	$[-1.1688, 2.66]$
	x_2	3.9037	$[0.23692, 7.5705]$	$[0.58791, 8.862]$	$[0.47297, 8.4462]$	$[-0.14766, 7.9551]$
$20, 0.025$	x_1	0.70573	$[-0.42784, 1.8393]$	$[0.030206, 2.3076]$	$[0.10206, 2.768]$	$[-0.85376, 2.2652]$
	x_2	4.1792	$[0.93936, 7.419]$	$[1.3748, 7.66]$	$[1.2585, 7.936]$	$[0.67262, 7.6858]$
$50, 0.01$	x_1	0.81365	$[-0.44374, 2.071]$	$[0.17395, 2.6482]$	$[0.12974, 3.0368]$	$[-0.89123, 2.5185]$
	x_2	3.7645	$[0.64616, 6.8828]$	$[0.73528, 6.9378]$	$[0.94692, 7.2942]$	$[0.40817, 7.1208]$
$100, 0.05$	x_1	0.86918	$[-0.40502, 2.1434]$	$[0.14043, 2.9166]$	$[0.14216, 3.2002]$	$[-0.82037, 2.5587]$
	x_2	3.6431	$[0.6853, 6.6009]$	$[0.77962, 7.0543]$	$[0.85571, 7.1334]$	$[0.48175, 6.8045]$

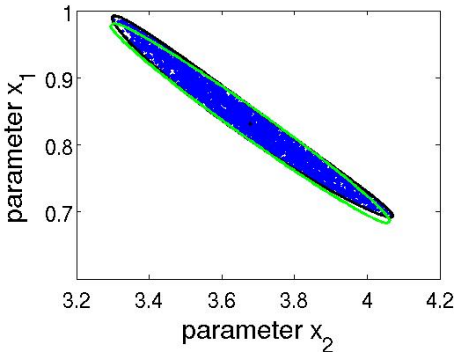

FIG. 5: $\sigma = 0.1$.

8. CONCLUSIONS

In this paper, we presented a new confidence region based on a second-order sensitivity analysis as well as a quadratic approximation of the covariance matrix. We analyzed and presented features of the introduced region. An important

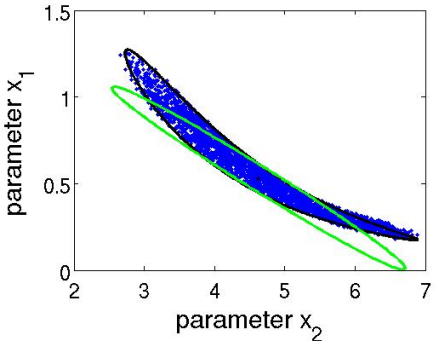

FIG. 6: $\sigma = 0.5$.

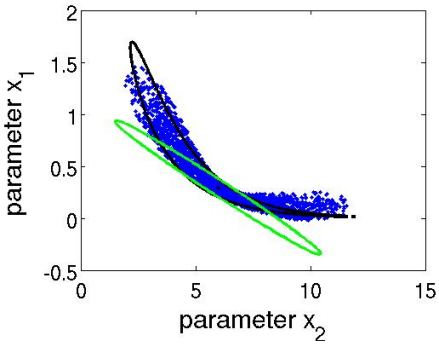

FIG. 7: $\sigma = 1$.

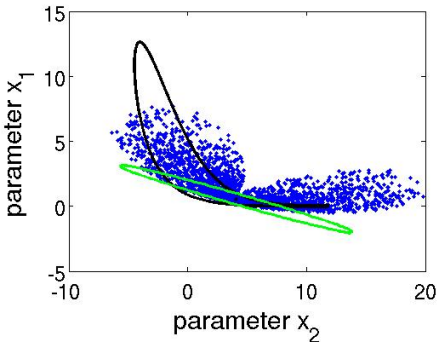

FIG. 8: $\sigma = 2$.

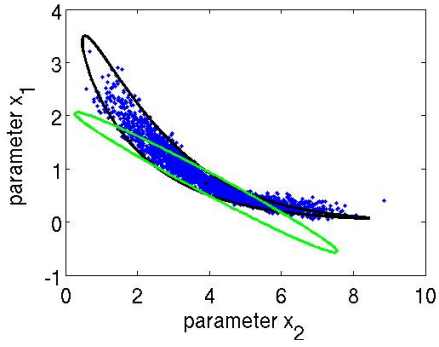

FIG. 9: $M = 10$.

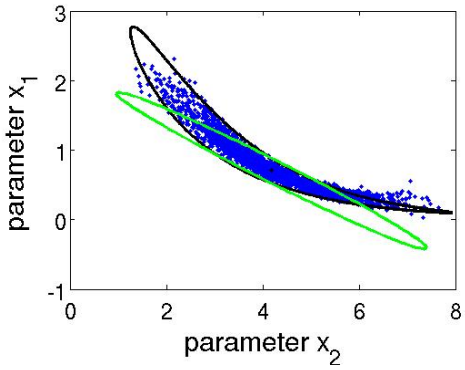

FIG. 10: $M = 20$.

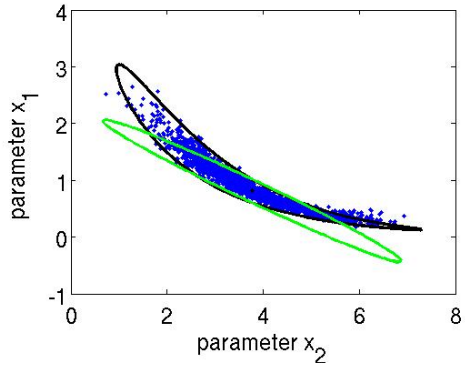

FIG. 11: $M = 50$.

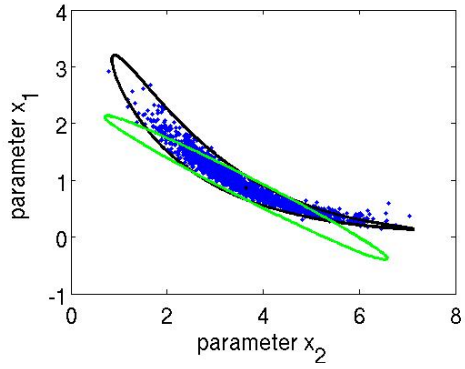

FIG. 12: $M = 100$.

result is that there is a strong analogy between the quadratic approximation and the Lipschitz constants κ and ω, which are also used to describe the local convergence rate of the Gauss-Newton method.

The features of the quadratic approximation of the confidence region are demonstrated by numerical examples. It is illustrated that the quadratic approximations of confidence regions are—in contrast to the linearized regions—very good approximations of the likelihood ratio confidence regions.

The results of the paper suggest that it is important to optimize the design of experiments not only based on the linear confidence analysis but to take into account the second-order information indicated by the Lipschitz constants ω and κ.

ACKNOWLEDGMENTS

This research was financed by BMBF-Program “Innovation in Mathematics for Industrial Services”. The authors are grateful to anonymous referees and guest editors of this issue, Alexey Chernov and Fabio Nobile, for their comments that allowed us to significantly improve the paper. We also would like to thank Hans Georg Bock for inspiring discussions and Hilke Stibbe for support in numerical computations.

REFERENCES

1. Bard, Y., *Nonlinear Parameter Estimation*, Academic Press, 1974.
2. Draper, N. R. and Smith, H., *Applied Regression Analysis*, Series in Probability and Mathematical Statistics, 2nd ed., Wiley, New York, 1981.
3. Gallant, A. R., *Nonlinear Statistical Models*, Wiley, New York, 1987.
4. Pázman, A., *Nonlinear Statistcal Models*, Kluwer Academic Publishers, Dordrecht, 1993.
5. Seber, G. A. F. and Wild, C. J., *Nonlinear Regression*, Wiley, New York, 1989.
6. Donaldson, J. and Schnabel, R., Computational experience with confidence regions and confidence intervals for non-linear least squares, *Technometrics*, 29(1):67–82, 1987.
7. Rooney, W. C. and Biegler, L. T., Design for model parameter uncertainty using nonlinear confidence regions, *AIChE J.*, 47(8):1794–1804, 2001.
8. Schwaab, M., Jr., E. C. B., Monteiro, J. L., and Pinto, J. C., Nonlinear parameter estimation through particle swarm optimization, *Chem. Eng. Sci*, 63(6):1542–1552, 2008.
9. Wiechert, W., Siefke, C., Graaf, A., and Marx, A., Bidirectional reaction steps in metabolic networks: II, *Flux Est. Stat. Anal. Biotech. Bioeng.*, 55(1):118–135, 1997.
10. Hamilton, D. C., Watts, D. G., and Bates, D. M., Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions, *Annals Stat.*, 10(2):386–393, 1982.
11. Potocký, R. and Ban, T. V., Confidence regions in nonlinear regression models, *Appl. Math.*, 37(1):29–39, 1992.
12. Dalai, M., Weyer, E., and Campi, M. C., Parameter identification for nonlinear systems: Guaranteed confidence regions through lscr, *Automatica*, 43(8):1418–1425, 2007.
13. Moselhy, T. A. E. and Marzouk, Y. M., Bayesian inference with optimal maps, *J. Comput. Phys.*, 231(23):7815–7858, 2012.
14. Bock, H.-G., Randwertproblemmethoden zur Parameteridentifizierung in Systemen nichtlinearer Differentialgleichungen, PhD thesis, Universitaet Bonn, Bonner Mathematische Schriften 183, 1987.
15. Nocedal, J. and Wright, S. J., *Numerical Optimization*, Springer Series in Operations Research, Springer-Verlag, New York, 1999.
16. Bock, H.-G., Kostina, E., and Schlöder, J., Numerical methods for parameter estimation in nonlinear differential algebraic equations, *GAMM-Mitteilungen*, 30(2):376–408, 2007.
17. Bock, H.-G., Kostina, E., and Kostyukova, O., Covariance matrices for parameter estimates of constrained parameter estimation problems, *SIAM J. Matrix Anal. Appl.*, 29(2):626–642, 2007.
18. Beale, E. M. L., Confidence regions in non-linear estimation, *J. Royal Stat. Soc. Series B*, 22(1):41–88, 1960.
19. Vanrolleghem, P. A. and Keesman, K. J., Identification of biodegradation models under model and data uncertainty, *Water Science and Technology*, 33(2):91–105, 1996.
20. Magnus, J. R. and Neudecker, H., *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, New York, 1999.
21. Marske, D., Biochemical oxygen demand data interpretation using sum of squares surface, PhD thesis, University of Wisconsin-Madison, 1967.
22. Bates, D. M. and Watts, D. G., *Nonlinear Regression Analysis and Its Applications*, Wiley, New York, 1988.
23. Daniel, C. and Wood, F. S., *Fitting Equations to Data*, Series in Probability and Mathematical Statistics, 2nd ed., Wiley, New York, 1980.
24. Keeping, E. S., *Introduction to Statistical Inference*, Van Nostrand Company, Princeton, 1962.